Trusted Digital Repository

Standards for Establishing Trusted Repositories for USGS Digital Assets

Introduction

These standards are intended to assist in selecting, specifying, building, operating, or enhancing trusted repositories for USGS digital scientific assets. This document includes a table for use by Bureau scientists and management (in collaboration with information technology (IT) staff) in the technical evaluation of systems for preserving these digital assets. The table establishes the minimum USGS standards for a trusted digital repository (refer to Level Three in the table below). The standards in the table are based on material from the Library of Congress-sponsored National Digital Stewardship Alliance (National Digital Stewardship Alliance, 2013). These standards do not cover physical data or address topics such as preservation policies, funding, or organizational competency and longevity, which are critical for data preservation but beyond the scope of this document.

For purposes of this document, important definitions related to preservation of USGS digital assets are as follows:

- **Long-term:** A period of time long enough for there to be concern about the loss of integrity of digital information held in a repository, including deterioration of storage media, changing technologies, support for old and new media and data formats, and a changing user community. This period extends into the indefinite future.
- **Sustainable format:** The ability to access an electronic record throughout its lifecycle, regardless of the technology used to create it. A sustainable format is one that increases the likelihood of a record being accessible in the future.
- **Checksums:** A checksum is a short mathematical digest of a file, which changes if any bit in the file changes. Checksums are used to detect unexpected changes in file content. Federal agencies, including the USGS, should use the following National Institute of Standards and Technology (NIST) approved checksums for new systems: SHA–224, SHA–256, SHA–384, and SHA–512.MD5 and SHA–1 checksums are widely used but not approved for new systems. For more information on checksums, refer to http://csrc.nist.gov/groups/ST/toolkit/secure_hashing.html.

Elements to Consider for Digital-Asset Preservation

When considering how to preserve digital assets you should address the following technical elements, for which standards are provided in the table below:

- Storage and Geographic Location – Storage systems, locations, and data duplication to prevent loss.
- Data Integrity – Procedures to prevent, detect, and recover from unexpected or deliberate changes.
- Information Security – Procedures to prevent human-caused corruption, deletion, and unauthorized access.
- Metadata – Documentation to enable contextual understanding and long-term usability.
- File Formats – File types, structures, and naming conventions to aid long-term preservation and reuse.
- Physical Media – Basic recommendations to reduce obsolescence risks that can threaten the readability of physical media.

Levels of Digital-Asset Preservation

There are four levels of digital asset preservation:

- **Level One**: Level One is the minimum criteria and activities needed to maintain digital assets through the life of a research project.
- **Level Two**: To continue improving upon repository functionality, implement Level Two elements after all Level One elements are in place.
- **Level Three**: Implement Level Three elements after all Level Two elements are in place. This is the USGS trusted digital repository minimum criteria for all long-term preservation records.
- **Level Four**: Level Four is the optimum level for which USGS should strive.

The Levels of Digital Preservation table below is based on a left-to-right progression. For each element, the columns describe four levels of increasing assurance for digital assets to be preserved. Additional guidelines are as follows:

- Each level adds requirements to the previous levels.
- To enhance an existing digital data repository, upgrade all elements to the same level.
- To achieve designation as a trusted digital repository, the repository must meet at least Level Three.
- For highest assurance of data preservation, specify all elements at Level Four.

Levels of Digital Preservation				
ELEMENT	**LEVEL ONE**	**LEVEL TWO**	**LEVEL THREE**	**LEVEL FOUR**

Storage and Geographic Location	• Two copies stored physically separate from each other • Transfer the digital content from temporary media into an established storage system • Managed storage system in place	• Three copies stored physically separate from each other • At least one copy in a different geographic location (off-site locations must follow NA RA 1571 guidelines) • Document the storage system and storage media	• At least one copy in a geographic location with a different disaster threat (e.g. hurricane-prone area versus an earthquake-prone area) • Maintain an obsolescence monitoring process for the storage system and media	• At least three copies in geographic locations with different disaster threats • Implement a comprehensive plan that keeps files and metadata on currently accessible systems and media
Data Integrity	• Verify checksums on ingest, if provided • Create checksums if not provided • Virus check all content	• Verify checksums on all data ingest • Use read-only procedures when working with original media	• Verify checksums at fixed interval of 2 years • Maintain logs of checksums and supply audit information on demand • Maintain procedures to detect corrupt data	• Verify checksums of all content in response to specific events or activities • Maintain procedures to replace or repair corrupted data • Ensure no one person has write access to all copies • Create, store, and verify a second, different checksum for all content
Information Security	• Identify who has authorization to read, write, move, and delete individual files • Limit authorizations to individual files	• Document access restrictions for content	• Maintain logs of who performed what actions on files, including deletions and preservation actions	• Perform audit of logs
Metadata	• Inventory of content and its storage location • Ensure backup and physical separation of inventory information • Adhere to current USGS metadata standards	• Store all relevant database management information • Store information describing changes to the structure or format of the data, including time of occurrence • Provide access to all forms of the metadata	• Preserve standard *technical*,*descriptive*, and *preservation* metadata	• Same as Level Three
File Formats	• Encourage the use of a limited set of documented and open file formats, codecs, compression schemes, and encapsulation schemes	• Inventory the file formats in use	• Monitor file format obsolescence issues	• Perform format migrations
Physical Media	• Inventory all physical media utilized including hard disks.	• Develop a plan to utilize trade studies to evaluate medias suitable for USGS purposes. • Begin to transition away from all media utilized that are 10 years or more in age.	• All non-recommended media have been properly disposed of following transition activities.	• Base all media choices on trade studies • All information is migrated from an older media to a newer media every 3 to 5 years, including hard disks.
Derived from Library of Congress, National Digital Stewardship Alliance, NDSA Levels of Digital Preservation: Version 1, February 2013.				

Roles and Responsibilities

- A repository manager or project chief ensures that all the table elements are addressed, although others, such as data managers or IT specialists, may be responsible for implementation and operation activities.
- Scientists and research staff will use the table criteria to recommend the suitability of a potential repository for preserving digital assets.
- Management officials will use the table criteria for reviewing and approving the selection of trusted digital repositories.
- In consultation with USGS scientist and managers, IT staff will use the table criteria for building, enhancing, or operating trusted digital repositories.

Additional Information

Additional information on preservation of USGS digital assets can be found at http://www.usgs.gov/datamanagement/preserve.php.